CSE574
Introduction
to Machine
Learning

Jue Guo

Parameters
vs. Hyperpa-
rameters

Classification
vs.
Regression

Model-Based
vs Instance-
Based
Learning

Shallow vs.
Deep
Learning

CSE574 Introduction to Machine Learning
Machine Learning: Notation and Definition Part 2

Jue Guo

University at Buffalo

January 27, 2024

Outline

- A **hyperparameter** is a property of a learning algorithm, usually (but not always) having a numerical value. That value influences the way the algorithm works. Hyperparameters aren't learned by the algorithm itself from data. They have to be set by the data analyst before running the algorithm.

- **Parameters** are variables that define the model learned by the learning algorithm. Parameters are directly modified by the learning algorithm based on the training data. The goal of learning is to find such values of parameters that make the model optimal in a certain sense.

Classification is a problem of automatically assigning a **label** to an **unlabeled example**. Spam detection is a famous example of classification.

- In machine learning, the classification problem is solved by a **classification learning algorithm** that takes a collection of **labeled examples** as inputs and produces a **model** that can take an unlabeled example as input and either directly output a label or output a number that can be used by the analyst to deduce the label. An example of such a number is a probability.

- In a classification problem, a label is a member of a finite set of **classes**. If the size of the set of classes is two ("sick"/"healthy", "spam"/"not spam"), we talk about **binary classification** (also called **binomial** in some sources). **Multiclass classification** (also called **multinomial**) is a classification problem with three or more classes.

- While some learning algorithms naturally allow for more than two classes, others are by nature binary classification algorithms. There are strategies allowing to turn a binary classification learning algorithm into a multiclass one. More on this later.

CSE574
Introduction
to Machine
Learning

Jue Guo

Parameters
vs. Hyperpa-
rameters

Classification
vs.
Regression

Model-Based
vs Instance-
Based
Learning

Shallow vs.
Deep
Learning

Regression is a problem of predicting a real-valued label (often called a **target**) given an unlabeled example.

- Estimating house price valuation based on house features, such as area, the number of bedrooms, location and so on is a famous example of regression.
- The regression problem is solved by a **regression learning algorithm** that takes a collection of labeled examples as inputs and produces a model that can take an unlabeled example as input and output a target.

Most supervised learning algorithms are model-based. We have already seen one such algorithm: SVM.

- **Model-based learning algorithms** use the training data to create a model that has **parameters** learned from the training data.
- In SVM, the two parameters we saw were $\mathbf{w}^*$ and b^*. After the model was built, the training data can be discarded.

Instance-based learning algorithms use the whole dataset as the model. One instance-based algorithm frequently used in practice is **k-Nearest Neighbors** (kNN).

- In classification, to predict a label for an input example the kNN algorithm looks at the close neighborhood of the input example in the space of feature vectors and outputs the label that it saw the most often in this close neighborhood.

A **shallow learning** algorithm learns the parameters of the model directly from the features of the training examples. Most supervised learning algorithms are shallow. The notorious exceptions are **neural network** learning algorithms, specifically those that build neural networks with more than one **layer** between input and output.

- Such neural networks are called **deep neural networks**. In **deep neural network** learning (or, simply, deep learning), contrary to shallow learning, most model parameters are learned not directly from the features of the training examples, but from the outputs of the preceding layers.

Don't worry if you don't understand what that means right now. We look at neural networks more closely in later section.

Questions?